

## ABSTRACT

In one embodiment of the invention, computerized methods are provided for analyzing transcript sequence clusters by aligning the transcripts with genomic sequences to determine whether a cluster needs to be sub-clustered and whether clusters should be merged. In addition, in some embodiments, transcript sequences are trimmed according to their alignment with their corresponding genomic sequences. The modified clusters and trimmed sequences may be used for nucleic acid probe array design.